

COMPUTER METHOD AND APPARATUS FOR COLLECTING PEOPLE
AND ORGANIZATION INFORMATION FROM WEB SITES

ABSTRACT OF THE DISCLOSURE

Computer processing means and method for searching and retrieving Web pages
5 to collect people and organization information are disclosed. A Web site of potential
interest is accessed. A subset of Web pages from the accessed site are determined for
processing. According to types of contents found on a subject Web page, extraction of
people and organization information is enabled. Internal links of a Web site are
collected and recorded in a links-to-visit table. To avoid duplicate processing of Web
10 sites, unique identifiers or Web site signatures are utilized. Respective time thresholds
(time-outs) for processing a Web site and for processing a Web page are employed. A
database is maintained for storing indications of domain URL's, names of respective
owners of the URL's as identified from the corresponding Web sites, type of each Web
site, processing frequencies, dates of last processings, outcomes of last processings, size
15 of each domain and number of data items found in last processing of each Web site.